Pharmacogenomics: accessing important alleles by imputation from commercial genome-wide SNP arrays

R. Liboredo[1] and S.D.J. Pena[1,2,3]

[1]Laboratório de Genômica Clínica, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
[2]Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
[3]GENE - Núcleo de Genética Médica de Minas Gerais, Belo Horizonte, MG, Brasil

Corresponding author: S.D.J. Pena
E-mail: spena@dcc.ufmg.br

ABSTRACT. Personalized medicine is becoming a medical reality, as important genotype-phenotype relationships are being unraveled. The availability of pharmacogenomic data is a key element of individualized care. In this study, we explored genotype imputation as a means to infer important pharmacogenomic alleles from a regular commercially available genome-wide SNP array. Using these arrays as a starting point can reduce testing costs, increasing access to these pharmacogenomic data and still retain a larger amount of genome-wide information. IMPUTE2 and MaCH-Admix were used to perform genotype imputation with a dense reference panel from 1000 Genomes data. We were able to correctly infer genotypes for the warfarin-related loci *VKORC1* and *CYP2C9* alleles *2, *3, *5, and *11 and also clopidogrel-related *CYP2C19* alleles *2 and *17 for a small sample of Brazilian individuals, as well as for HapMap samples. The success of

an imputation approach in admixed samples using publicly available reference panels can encourage further imputation initiatives in those populations.

Key words: Pharmacogenomics; Imputation; Warfarin; Clopidogrel; Personalized medicine

INTRODUCTION

Since the release of the complete sequence of the human genome, there is a growing interest in estimating the influence of genomic variation in the predisposition to disease and individual response to drugs (Cavalli-Sforza, 2005). Several studies in that field have pointed towards an increasingly individualized approach to healthcare. Some of the most important aspects of personalized medicine are pharmacogenomic tests for several loci of well-known gene-drug relationships.

However, even for a single drug, we can expect more than one associated gene, thus requiring a multigenic approach (Johnson et al., 2013). Testing multiple loci is time consuming and has a low cost-effectiveness ratio. There are single nucleotide polymorphism (SNP) arrays specifically designed to access pharmacogenomic information at a reasonable price. However, their price is considerable because of the small market. Moreover, once the pharmacogenomic panel has been customized, any change introduced into the gene content is complex and costly. On the other hand, there are commercially available genome-wide SNP arrays that have a lower cost and provide further information on the most common genetic variants throughout the genome. Nevertheless, they do not target specific genes and, as such, cannot be used directly to assess exonic pharmacogenetic polymorphisms.

Genotype imputation has been widely used in genome-wide association studies (GWAS), usually being performed across the whole genome as an important step to increase the power of GWAS and also in smaller regions to guide fine-mapping studies (Jakobsson et al., 2008; Nothnagel et al., 2009; Marchini and Howie, 2010; Howie et al., 2012; Lu et al., 2012). It is a statistical way to use known haplotypes of a reference panel to infer missing genotypes in a sample of individuals initially genotyped at a subset of that panel of SNPs. Imputation works by making use of linkage disequilibrium, finding haplotype segments that are shared by both the panel and study individuals and using them to fill in missing data (Kong et al., 2008; Marchini and Howie, 2010). Imputation of biallelic pharmacogenomic variants has been reported as a valid method (Mijatovic et al., 2012), but little methodological research exists for imputation in admixed populations (Liu et al., 2013).

The aim of this study was to ascertain whether genotype imputation could be used to obtain individual pharmacogenomic information from commercial genome-wide SPN array data. As models, we decided to use warfarin and clopidogrel metabolic variants in a small sample of Brazilian individuals, using HapMap samples as imputation controls.

Introduced in the 1950s, warfarin is a coumarin anticoagulant widely used to prevent thromboembolic events (Lee and Klein, 2013). Prescribing the right dose is important to assure efficacy and safety of the drug, since the therapeutic window is narrow. Moreover, inter-individual differences in dose requirements are a concern, since a large genetic and environmental variation exists in the optimal dose (Perini et al., 2008; Takeuchi et al., 2009). Among the relevant genetic factors involved, two genes are especially known to participate: *CYP2C9*

(cytochrome P450, family 2, subfamily C, polypeptide 9) and *VKORC1* (vitamin K epoxide reductase complex, subunit 1).

We describe here the imputation of four alleles in *CYP2C9*: *2 (tagged by rs1799853), *3 (tagged by rs1057910), *5 (rs28371686), and *11 (rs28371685). The two minor alleles *2 and *3 are well known to be associated with reduced warfarin metabolism and thus are very clinically relevant. The *CYP2C9* alleles *5 and *11 occur mostly in African populations, and therefore are commonly considered in pharmacogenomic studies in the Brazilian population (Suarez-Kurtz, 2010). In the *VKORC1* locus, we imputed the noncoding variant rs923231, previously reported to be the definer of the warfarin sensitive haplotype, found to be the best predictor of warfarin dose (Takeuchi et al., 2009; Lee and Klein, 2013). These well-known genotypes at *CYP2C9* and *VKORC1* have led to a drug label dosage recommendation from the U.S. Food and Drug Administration (FDA), and to the development of several dosing algorithms that take them into account, showing a real possibility of translating pharmacogenomics research into clinical practice (Lee and Klein, 2013).

Clopidogrel is an anticoagulant that needs to be converted into its active metabolite to exert its antiplatelet effect. The gene product of *CYP2C19* (cytochrome P450, family 2, subfamily C, polypeptide 19) mediates this conversion. The most important alleles associated with clopidogrel are *2 (tag SNP rs4244285) and *17 (tag SNP rs12248560) (Harmsze et al., 2012) and both were imputed in this study. The variant allele *2 is associated with decreased platelet response to clopidogrel treatment (Hulot et al., 2006; Maeda et al., 2011) and the *17 allele is associated with an exaggerated sensitivity to treatment, with enhanced risk of bleeding (Harmsze et al., 2012).

MATERIAL AND METHODS

Samples

The commercially available Genome-Wide Human SNP Array 6.0 (Affy 6.0) from Affymetrix was used for this study. Affy 6.0 genotype results for 13 Brazilian individuals were kindly made available by GENE - Núcleo de Genética Médica (http://www.laboratoriogene.com.br/). Genotype results for 270 HapMap1 individuals, including European descendants (CEU) as well as African Yoruba (YRI), Han Chinese (CHB) and Japanese (JPT) populations, also genotyped with the Affy 6.0 array, were obtained from Affymetrix (The International HapMap Consortium, 2005).

As reference panels, we used all available data, independent of population of origin, from the 1000 Genomes Project (1000 Genomes) (Abecasis et al., 2010, 2012). This dispensed with the need of dealing with the complex choice of achieving the right combination of reference panels to mimic admixed populations, increasing imputation accuracy and even making it faster (Howie et al., 2011). Improved imputation algorithms can handle these large panels and account for admixture (Howie et al., 2009; Marchini and Howie, 2010; Li et al., 2010; Liu et al., 2013).

Input data management and pre-imputation quality control

Affymetrix Genotyping Console (GC) software version 4.1.2.837 was used to read,

filter and export the Affy 6.0 genotype results. PLINK was used to filter out SNPs with low call rate (below 95%), minor allele frequency (MAF) lower than 1%, and SNPs that departed from Hardy Weinberg equilibrium ($P < 10^{-6}$). Scripts were written in Python to adapt and convert genotyping data formats to fit either IMPUTE2 or MaCH-Admix. Gtool was also used to convert .ped files into .gen format.

Imputation

IMPUTE2 and MaCH-Admix are freely available programs for academic use and were downloaded from the IMPUTE2 (http://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html) and MaCH-Admix (http://www.sph.umich.edu/csg/yli/mach/download/1000G.2012-03-14.html) websites, together with the reference panels and recombination rate files. Strand alignment information files were downloaded from the Wellcome Trust Centre for Human Genetics in Oxford website (http://www.well.ox.ac.uk/~wrayner/strand/). The reference panel version used was the Phase 1 integrated from 1000 genomes, containing 2184 haplotypes and approximately 39 million variants, including SNPs, short insertion-deletion polymorphisms (Indels) and large deletions.

Default parameters of imputation programs were altered to maximize accuracy. An increase in the number of haplotypes used to perform the phasing step (at IMPUTE2 -k from default 80 to 200 and at MaCH-Admix -states from default 100 to 200) and the number of haplotypes used to perform imputation (at IMPUTE2 -k_hap from default 500 to 1000 and corresponding MaCH-Admix -imputeStates from default 500 to 1000) were the most relevant changes. Imputation was performed in chunks of 5 Mb, with the target SNP preferably placed at the center of that region. A pre-phasing step with Shapeit prior to imputation was not used because, although it reduces the running time and computational costs, it can also slightly decrease the accuracy (Howie et al., 2012). Scripts in Python were written to search and extract the genotypes of the SNPs of interest from the imputation output files. A post-imputation calling threshold was set within these scripts to filter IMPUTE2 results. Genotypes that did not meet that threshold were reported as non-informative (NI).

Comparison with experimental genotyping

Molecular genotyping was performed for the 13 Brazilian samples for comparison with imputed genotypes at *CYP2C9* and *VKORC1*. Polymerase chain reaction (PCR) conditions were: 95°C for 5 min, 5 cycles of 1 min at 95°C, 1 min at 58°C and 1 min at 72°C, followed by 25 cycles of 1 min at 95°C, 1 min at 55°C and 1 min at 72°C, and a final extension step of 30 min at 72°C. Primer sequences are listed in Table 1. In each batch of samples, controls of known genotype and blank tubes with no DNA were included. *CYP2C9*3* was genotyped using two different reactions for each sample, one for each possible genotype [ARMS-PCR (Newton et al., 1989)]. For *CYP2C9*2* and *VKORC1*, PCR products were digested at 37°C for 24 h with the restriction enzymes *Ava*II and *Msp*I, respectively, resulting in one or two fragments, depending on the sample genotype (Sconce et al., 2005). The results were visualized by polyacrylamide gel electrophoresis with silver staining. Whole exome sequencing (WES) data were also available for 7 of the 13 samples provided by GENE (data not shown).

Table 1. PCR primer sequences for *CYP2C9**2, *CYP2C9**3, and *VKORC1*.

Gene/allele	Primer
CYP2C9*2 F	5'-TAGTTTCGTTTCTCTTCCTGTTAG-3'
CYP2C9*2 R	5'-GAGAAGATAGTAGTCCAGTAAGGT-3'
CYP2C9*3 F	5'-CTGAATTGCTACAACAAATGTG-3'
CYP2C9*3 R nor	5'-TGGTGGGGAGAAGGTCTAT-3'
CYP2C9*3 R mut	5'-TGGTGGGGAGAAGGTCTAG-3'
VKORC1 F	5'-GCCAGCAGGAGAGGGAAATA-3'
VKORC1 R	5'-AGTTTGGACTACAGGTGCCT-3'

For the 270 HapMap samples, genotyping results for all studied SNPs were downloaded from Ensembl database (http://www.ensembl.org/Homo_sapiens/Info/Index), as well as from the NCBI dbSNP website (http://www.ncbi.nlm.nih.gov/projects/SNP/). The number of genotypes available varied according to the SNP studied.

RESULTS

Experimental genotypes for all 13 Brazilian samples were successfully obtained for *CYP2C9* **2*, **3* and *VKORC1* variants and can be seen in Table 2. Internal imputation quality estimates were high for all SNPs tested (over 0.85) and similar for both available computer programs. IMPUTE2 performed slightly better than MaCH-Admix, on average (Table 3). Genotype imputation results were compared to experimental genotypes confirming the accuracy predicted by these estimates. Those comparisons can be seen for experimental and IMPUTE2 results in **Table S1**.

Table 2. Experimental genotypes obtained for the 13 Brazilian samples.

Individual	*VKORC1*	*CYP2C9**2	*CYP2C9**3
FAM1 son	C T	T T	A A
FAM1 father	T T	C T	A A
FAM1 mother	C C	T T	A A
FAM2 son	C C	C T	A A
FAM2 mother	C C	C C	A A
FAM3 mother	C T	C T	A A
FAM3 grandmother	C C	C T	A A
FAM3 grandfather	C T	C C	A A
FAM3 son	C T	C T	A A
IND 4	C T	C T	A A
IND 5	C T	C C	A A
IND 6	C T	C C	A A
IND 7	C T	C C	A A

Table 3. IMPUTE2 and MaCH-Admix internal quality estimates for imputation of the loci studied.

Gene/allele	SNP	IMPUTE2	MaCH-Admix
CYP2C9*2	rs1799853	0.999	0.989
CYP2C9*3	rs1057910	0.961	0.984
CYP2C9*5	rs28371686	0.927	0.863
CYP2C9*11	rs28371685	0.946	0.874
VKORC1	rs9923231	0.999	0.993
CYP2C19*2	rs4244285	1.000	0.980
CYP2C19*17	rs12248560	0.982	0.959

Imputation of *CYP2C9* genotypes

For the *CYP2C9*2* allele (rs1799853 C>T), all the 13 Brazilian genotypes were also correctly inferred by IMPUTE2 (there was one mismatch at the MaCH-Admix output). In total, 257 of the 258 available HapMap genotypes were correctly inferred by imputation, with one single mismatch occurring at both IMPUTE2 and MaCH-Admix at the sample NA10854 (CEU). That genotype was imputed as a homozygous variant (TT), while the available experimental genotype from HapMap was reported as heterozygous (CT). Only two other samples (NA11830 and NA11831, both from CEU) showed the homozygous variant genotype for that SNP. We noticed that for those two samples, there were two conflicting experimental genotypes available, one from the 1000G (TT) and another from HapMap (CT). In view of this ambiguity, we decided to take a closer look at our data to see if these 3 samples could be sharing the same haplotype. Using an optional parameter of IMPUTE2 (-phase), we were able to access the presumed haplotypes generated by the phasing of the Affy6.0 genotype results. We noticed an identical region from position 96386206 to position 96749181 (Chr10) for the three samples, suggesting that they indeed shared the same haplotype at the tag SNP region. With that result, and considering that we used the 1000G results as our reference panel, we confirmed the imputation call for NA10854 as TT.

For the *CYP2C9*3* (rs1057910 A>C), all 270 HapMap genotypes were available and correctly inferred by MaCH-Admix. One genotype did not meet the IMPUTE2 post-imputation calling threshold, even when it was reduced to 70%, therefore being reported as NI. The other 269 samples had their genotypes correctly inferred, and 11 of the 13 Brazilian genotypes were correctly inferred by both software. Two genotypes from two related samples (father and daughter) did not meet the IMPUTE2 post-imputation calling threshold and were reported as NI. However, by lowering IMPUTE2 threshold to 70%, the imputed genotypes matched the experimental genotypes (AA). The same two genotypes were incorrectly inferred by MaCH-Admix as heterozygous (AC).

At the tag SNP for the *CYP2C9*5* variant (rs28371686 C>G), 153 genotypes were available for HapMap samples. All 153 were correctly imputed by MaCH-Admix and IMPUTE2, including the only one that showed the variant allele, the YRI sample NA18853. Experimental genotypes from exome sequencing were available for 7 of the 13 Brazilian samples, and corroborated the imputed data. Two related samples (mother and son) from the 13 Brazilian had one variant (CG genotype); however, experimental genotypes were not available for those individuals.

Individual genotypes were available for 159 HapMap samples at the variant *CYP2C9*11* (rs28371685 C>T) and all of them were correctly imputed by IMPUTE2 and MaCH-Admix. Only two YRI samples were heterozygous for this variant, and the other 157 were homozygous CC. Experimental genotypes from exome sequencing was available for 7 of the 13 Brazilian samples and corroborated the imputed data.

Imputation of *VKORC1* genotypes

A summary of IMPUTE2 results for HapMap samples at warfarin-related alleles and their overall genotype distribution are shown in Table 4. Combined results of *VKORC1* genotypes and *CYP2C9* diplotypes imputed for Brazilian samples are shown in Table 5, including

imputation results for *5 and *11, although they were not experimentally confirmed. Diplotype distribution of *CYP2C9* alleles can be seen in **Table S2**.

Table 4. Summary of imputation results and genotype distribution with IMPUTE2 for warfarin related alleles on HapMap samples (0 indicates the ancestral allele and 1 indicates the variant allele).

Allele	HapMap imputed genotypes					
	N	0/0	0/1	1/1	NI	Errors
CYP2C9*2	258	243	12	3	0	0
CYP2C9*3	270	253	16	0	1	0
CYP2C9*5	153	152	1	0	0	0
CYP2C9*11	159	157	2	0	0	0
VKORC1	262	117	57	88	0	0

Table 5. *CYP2C9* diplotypes and *VKORC1* genotypes imputed for Brazilian samples.

Individual	*VKORC1*	*CYP2C9*
FAM1 son	C T	*2/*2
FAM1 father	T T	*1/*2
FAM1 mother	C C	*2/*2
FAM2 son	C C	*2/*5
FAM2 mother	C C	*1/*5
FAM3 mother	C T	*1/*2
FAM3 grandmother	C C	*1/*2
FAM3 grandfather	C T	*1/*1
FAM3 son	C T	*1/*2
IND 4	C T	*1/*2
IND 5	C T	*1/*1
IND 6	C T	*1/*1
IND 7	C T	*1/*1

For *VKORC1* rs9923231, experimental genotypes were available for 262 HapMap samples and all 13 the Brazilian samples. Comparison showed that all samples had their genotypes correctly inferred by IMPUTE2 and MaCH-Admix, with a 95% post-imputation calling threshold. The most common genotype found for the Brazilian sample was the heterozygous CT, which was also the most common for the HapMap CEU sample. The homozygous CC was the most common genotype found in the YRI sample, whereas the homozygous TT was the most common for the CHB and JPT samples combined. Genotype distribution of rs9923231 for all samples can be seen in **Table S3**.

Imputation of *CYP2C19* genotypes

For the clopidogrel-related allele *CYP2C19**2 (rs4244285 G>A), all 266 HapMap available genotypes were correctly inferred by both IMPUTE2 and MaCH-Admix. Seven of the 13 Brazilian samples had available exome results that corroborated the imputation results. A summary of the genotype distribution of *CYP2C19**2 and *17 can be seen in **Table S4**.

For *CYP2C19**17 (rs12248560 C>T), all the 270 available HapMap genotypes were correctly inferred by MaCH-Admix. IMPUTE2 results with a post-imputation calling thresh-

old set to 90% returned 266 correct genotypes and 4 NI genotypes. By lowering the threshold to 85%, we obtained 2 NI, and with 75%, we had all imputed genotypes correctly inferred. For the Brazilian samples, 2 returned as NI genotypes even with the threshold set to 75%. It was not possible to compare imputation and exome sequencing results as the tag SNP for the allele rs12248560 is not in an exon.

Considering the combined results for the *CYP2C19* alleles studied, the most common *CYP2C19* diplotype for HapMap and Brazilian samples was *1/*1, and the second most common was *1/*17 for Brazilian, CEU and YRI samples, while the *1/*2 diplotype was the second most common for combined CHB and JPT samples. The numeric distribution can be seen in Table 6.

Table 6. Diplotype distribution of *CYP2C19* alleles studied.

Diplotype	CYP2C19			
	BRA	CEU	CHB+JPT	YRI
*1/*1	5	36	44	30
*1/*2	2	15	40	12
*1/*17	3	27	0	29
*2/*17	1	5	2	8
*2/*2	0	3	3	3
*17/*17	0	2	0	7
N	11	88	89	89

DISCUSSION

Our results showed that it was possible, using currently available reference panels, to obtain reliable genotypes from genotype imputation on the loci studied, not only for populations that are well represented on the reference panel, but also for samples from an admixed population. Even rare alleles that occur in one parental population but not the others, such as *CYP2C9*5* and *11, were correctly imputed. The two imputation software used, IMPUTE2 and MaCH-Admix, showed similar results.

Using genotype imputation, we were able to access rare genotypes that are often not included on common tests, such as *CYP2C9*5*. Regular tests are aimed to identify the two most common alleles (*2 and*3), but in admixed populations, it is important to widen that search. In our small Brazilian sample, two individuals showed a rare variant allele.

Using existing databases to successfully perform imputation for admixed populations is a reality, but population-specific panels are expected to become available soon. We hope that the use of these panels combined with the large 1000 Genomes panel may further improve imputation accuracy.

ACKNOWLEDGMENTS

Supplementary material

REFERENCES

Abecasis GR, Altshuler D, Auton A, Brooks LD, et al. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467: 1061-1073.

Abecasis GR, Auton A, Brooks LD, DePristo MA, et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56-65.

Cavalli-Sforza LL (2005). The Human Genome Diversity Project: past, present and future. *Nat. Rev. Genet.* 6: 333-340.

Harmsze AM, van Werkum JW, Hackeng CM, Ruven HJ, et al. (2012). The influence of CYP2C19*2 and *17 on on-treatment platelet reactivity and bleeding events in patients undergoing elective coronary stenting. *Pharmacogenet. Genomics* 22: 169-175.

Howie BN, Donnelly P and Marchini J (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5: e1000529.

Howie B, Marchini J and Stephens M (2011). Genotype imputation with thousands of genomes. *G3* (*Bethesda*) 1: 457-470.

Howie B, Fuchsberger C, Stephens M, Marchini J, et al. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* 44: 955-959.

Hulot JS, Bura A, Villard E, Azizi M, et al. (2006). Cytochrome P450 2C19 loss-of-function polymorphism is a major determinant of clopidogrel responsiveness in healthy subjects. *Blood* 108: 2244-2247.

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, et al. (2008). Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451: 998-1003.

Johnson JA, Klein TE and Relling MV (2013). Clinical implementation of pharmacogenetics: more than one gene at a time. *Clin. Pharmacol. Ther.* 93: 384-385.

Kong A, Masson G, Frigge ML, Gylfason A, et al. (2008). Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* 40: 1068-1075.

Lee MT and Klein TE (2013). Pharmacogenetics of warfarin: challenges and opportunities. *J. Hum. Genet.* 58: 334-338.

Li Y, Willer CJ, Ding J, Scheet P, et al. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* 34: 816-834.

Liu EY, Li M, Wang W and Li Y (2013). MaCH-admix: genotype imputation for admixed populations. *Genet. Epidemiol.* 37: 25-37.

Lu JT, Wang Y, Gibbs RA and Yu F (2012). Characterizing linkage disequilibrium and evaluating imputation power of human genomic insertion-deletion polymorphisms. *Genome Biol.* 13: R15.

Maeda A, Ando H, Asai T, Ishiguro H, et al. (2011). Differential impacts of CYP2C19 gene polymorphisms on the antiplatelet effects of clopidogrel and ticlopidine. *Clin. Pharmacol. Ther.* 89: 229-233.

Marchini J and Howie B (2010). Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* 11: 499-511.

Mijatovic V, Iacobucci I, Sazzini M, Xumerle L, et al. (2012). Imputation reliability on DNA biallelic markers for drug metabolism studies. *BMC Bioinformatics* 13 (Suppl 14): S7.

Newton CR, Graham A, Heptinstall LE, Powell SJ, et al. (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* 17: 2503-2516.

Nothnagel M, Ellinghaus D, Schreiber S, Krawczak M, et al. (2009). A comprehensive evaluation of SNP genotype imputation. *Hum. Genet.* 125: 163-171.

Perini JA, Struchiner CJ, Silva-Assuncao E, Santana IS, et al. (2008). Pharmacogenetics of warfarin: development of a dosing algorithm for Brazilian patients. *Clin. Pharmacol. Ther.* 84: 722-728.

Sconce EA, Khan TI, Wynne HA, Avery P, et al. (2005). The impact of CYP2C9 and VKORC1 genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. *Blood* 106: 2329-2333.

Suarez-Kurtz G (2010). Pharmacogenetics in the Brazilian population. *Front Pharmacol.* 1: 118.

Takeuchi F, McGinnis R, Bourgeois S, Barnes C, et al. (2009). A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. *PLoS Genet.* 5: e1000433.

The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* 437: 1299-1320.